A neural analogue of the worst performance rule: Insights from single-trial event-related potentials

Saville, C.W.; Beckles, K.D.; MacLeod, C.A.; Feige, B.; Biscaldi, M.; Beauducel, A.; Klein, C.

Intelligence

DOI:
10.1016/j.intell.2015.12.005

Published: 31/12/2015

Peer reviewed version

Intelligence xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

A neural analogue of the worst performance rule: Insights from single-trial event-related potentials

Christopher W.N. Saville [a,b,*], Kevin D.O. Beckles [a], Catherine A. MacLeod [c], Bernd Feige [d], Monica Biscaldi [b], André Beauducel [e], Christoph Klein [a,b,f]

[a] School of Psychology, Bangor University, UK
[b] Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University of Freiburg, Germany
[c] Institute of Medical and Social Care Research, Bangor University, UK
[d] Department of Psychiatry and Psychotherapy, University of Freiburg, Germany
[e] Institute of Psychology, University of Bonn, Germany
[f] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Medical Faculty, University of Cologne, Germany

ARTICLE INFO

Article history:
Received 29 May 2015
Received in revised form 17 December 2015
Accepted 18 December 2015
Available online xxxx

Keywords:
P300
Intelligence
g
Reaction time variability
Mental chronometry

ABSTRACT

The worst performance rule is the tendency for participants' slowest reaction times to correlate more with psychometric intelligence than their faster reaction times. Reaction times, however, are influenced by the duration of multiple perceptual, attentional, and motor sub-processes, and it is unclear whether the same pattern exists in these sub-processes as well. We used single-trial event-related potentials to identify whether a worst performance rule pattern could be found in stimulus and response-locked P3b latency distributions and scores on a test of non-verbal psychometric intelligence.

Fifty participants carried out a set of working memory oddball tasks, while electroencephalographic data were collected, and the British Version of the Intelligence Structure Test, in a separate session. Single-trial P3b latencies were identified in stimulus and response-locked data and a novel quantile bootstrapping method was used to identify which quantiles of the P3b latency distributions correlated most with test scores.

In stimulus-locked data, correlations between quantile mean and test scores became more negative with increasing quantile, showing clear evidence of a worst performance rule pattern. In response-locked data, low scorers showed more extreme latencies in both tails of the distribution. However we did not observe a worst performance rule in behavioural data. These data suggest that psychometric intelligence is also associated with response-related processes, which may also contribute to the association between psychometric intelligence and reaction time variability.

1. Introduction

Reaction time (RT) is a long established correlate of psychometric intelligence (Beck, 1933; Jensen, 2006), with higher ability individuals able to carry out cognitive operations faster than those with lower ability. In order to obtain reliable RT measurements, it is a standard practice to run many trials, which are aggregated to obtain a single representative measure of speed. This procedure assumes that all intra-subject variability (ISV) is meaningless noise, to be minimised in order to estimate a 'true' RT. However it is becoming increasingly clear that intra-subject RT variability is not only a reliable trait in its own right Saville et al. (2011b), but also an independent correlate of psychometric intelligence (Deary, Der, & Ford, 2001; Jensen, 1992), and it is now widely accepted that higher *g* is associated with not just faster RTs, but less variable RTs as well.

A more specific link between RT and *g*, is the *worst performance rule*. Larson and Alderton (1990) suggested that measuring overall variability could conflate different sources of variance. By averaging RTs within quantile bands and correlating *g* with each quantile's mean separately, they found that the correlation between *g* and RT increased with increasing quantile, suggesting that it was the slowest RTs that were driving the relationship between ISV and *g*, a finding replicated a number of times (Coyle, 2003; Kranzler, 1992; but see Salthouse, 1998). The worst performance rule is a striking finding because of what it appears to suggest about intelligence: the difference between high and low *g* participants is not in their maximum capability, but in their ability to maintain this level over the course of a task. The speed-related facet of intelligence may be related to performance more than competence.

Despite the above suggestion that the worst performance rule reflects an association between *g* and sustained attention, this remains unsubstantiated. The existing literature, however, advances two plausible

* Corresponding author at: North Wales Clinical Psychology Programme, School of Psychology, Bangor University, Ffordd Penrallt, Bangor, Gwynedd, Wales LL57 2AS, UK.
E-mail address: c.saville@bangor.ac.uk (C.W.N. Saville).

http://dx.doi.org/10.1016/j.intell.2015.12.005
0160-2896/

models. One possibility is that the worst performance rule is a result of an association between *g* and sustained attention. Lapses in attention could be more common in participants with low *g*. Such lapses would cause delays to the start of stimulus-processing, and thereby lead to a greater incidence of slow RTs. Such lapses would, however, likely leave later response-processing stages largely unaffected, albeit delayed.

A second model is that RTs can be modelled as the accrual of information to decide between two possible responses, with a decision being made once a certain threshold has been reached. Such a conception of RTs appears to be a good fit for a variety of data, and the most famous of this class of models is the Diffusion Model (Ratcliff, 1978). Simulations have shown that this model can reproduce the worst performance rule by representing lower *g* as a slower rate of information accrual, or as a more stringent decision threshold (Ratcliff, Schmiedek, & McKoon, 2008). This explanation contrasts with the attentional lapse model in that slow RTs do not represent a distinct group from 'normal' RTs, but part of the same distribution.

Finally, a largely unconsidered possibility is that processes translating an abstract decision about the stimulus into a motor plan could contribute to the worst performance rule. Such an association could reflect a specific association between such processes and *g*, or they could reflect an association between *g* and a more general property of neurocognition, such as axonal myelination (Miller, 1994), and such a 'global' property of neural transmission could lead to variable delays at multiple stages of processing.

Single-trial event-related potentials (Saville et al., 2011a, Saville et al., 2014; 2015b) offer an opportunity to investigate this. By identifying neural markers of covert processing steps in each trial we can estimate the latency distributions of sub-processes that underlie an RT to identify whether each subcomponent displays a worst performance rule pattern. This will help to differentiate between the three models described above.

Event-related potentials (ERPs) typically require averaging to be measured reliably, but some have sufficient signal-to-noise ratio to be measured in single trials. One component which is amenable to single-trial analysis is the P3b (Gerson, Parra, & Sajda, 2005; Saville et al., 2011a). The P3b is thought to represent the completion of a neural decision (O'Connell, Dockree, & Kelly, 2012; Verleger, Jaśkowski, & Wascher, 2005), or the propagation of such a decision (Nieuwenhuis, Aston-Jones, & Cohen, 2005), before the decision is converted into a motor plan. P3b latency is thus sensitive to some, but not all of the same factors as RT (Kutas, McCarthy, & Donchin, 1977). An interesting property of the P3b is that it is time-locked to both stimulus-onset and response, in line with its putative role as a bridge between stimulus-processing and response execution (Verleger et al., 2005; Verleger, Schroll, & Hamker, 2013). P3b latency can thus be measured relative to both events, and has a different significance depending on the time-locking event, with stimulus-locked effects being consistent with an effect on pre-decision neurocognitive processes and response-locked effects indicating post-decision processes.

In the present study, we will carry out a worst performance rule analysis of single-trial P3b latency distributions. This analysis will be carried out on both stimulus and response-locked P3b latencies, in order to identify whether the worst performance rule pattern can be identified in these latencies. Evidence for the pattern in stimulus-locked data alone would suggest that this pattern is due to early processing factors, such as lapses in attention. Evidence for the pattern in response-locked data alone would suggest that response-planning is responsible. Evidence for a worst performance rule in both distributions would suggest that a single general factor, or multiple different factors, underlies this effect.

2. Methods

The study was approved by the Ethics and Research Governance Committee at the School of Psychology, Bangor University.

2.1. Participants

Fifty participants, primarily students from Bangor University, took part in the present study. Mean age was 22 (±3.6), 35 were female, and three were left-handed.

2.2. Intelligence measures

Participants completed the English language version of the Basic Module of the Intelligence Structure Test (Beauducel, Liepmann, Horn, & Brocke, 2010) on a different day to the EEG session, either alone or in small groups. The Intelligence Structure Test is substantially correlated with Cattell's Culture-Fair-Intelligence test (CFT-20, $r = .63$, $N = 180$) and Raven's Advanced Progressive Matrices ($r = .69$, $N = 244$), so represents a good measure of *g* (Beauducel et al., 2010). As some participants were not native speakers of English, we created a composite non-verbal intelligence score by running maximum likelihood factor analysis (as implemented by the *factanal* function in R [R Core Development Team, 2012]) on the numeric and figural subtests, and creating weighted sum scores for each participant using the factor weightings. The numeric subtests were as follows: Calculations (basic arithmetic), Number Series (identifying the next number in a sequence burying a mathematical rule), and Number Signs (choosing the correct mathematical operator to make an equation correct). The figural subtests were as follows: Figure Selection (choosing which shape could be formed from a set of components), Cubes (choosing which cube was a rotated version of an exemplar), and Matrices (akin to Raven's progressive matrices). Each subtest included 20 items and participants were given between 6 and 10 min, depending on the subtest in question. Full details can be found in Beauducel et al. (2010).

Numeric scores averaged 34.3 (±11.3) and figural scores averaged 33.5 (±7.1). The numeric and figural scores shared a great deal of variance ($r = .63$) but verbal scores were less highly correlated with both numeric ($r = .26$) and figural ($r = .22$) scores, suggesting that our exclusion of verbal scores was appropriate.

2.3. Apparatus and materials

Direct-current EEG data were recorded with 63 Ag/AgCl electrodes (Falk Minow, Germany) in a 10–10 montage (American Electroencephalographic Society, 1991) and two infra-orbital electrodes. Electrodes Cz and FPz were used as recording reference and ground respectively. Abralyt high chloride gel (Easycap, Germany) was used to reduce impedances to ≤5 kΩ at all electrodes before data collection. The signal was differentially-amplified by two BrainAmp DC amplifiers (0.1 μV resolution, 1 kHz sampling rate), and recorded with BrainVision Recorder (both Brain Products, Germany). Participants were tested in a sound-attenuated Faraday cage, and stimuli were presented on a 17″ LCD monitor with an electrically-shielded power source.

2.4. Stimuli and procedure

The EEG session involved three *n-back tasks*: a zero-back task (0BT), a one-back task (1BT), and a two-back task (2BT). In each, participants made speeded responses to sequences of letters. Participants responded with their right hand to *oddballs* (25% of trials) and their left to *standards* (75% of trials). In the 0BT, the letter 'E' was the oddball; in the 1BT, letters matching the previous letter were oddballs; in the 2BT, letters matching the letter before last were oddballs. So in the 0BT, the final letter in the sequence ABCE would be an oddball; in the 1BT the final letter in the sequence ABCC would be an oddball; and in the 2BT, the final letter in the sequence ABCB would be an oddball.

Participants were asked to respond as quickly and accurately as possible to all trials by manually pressing keys on a computer keyboard. We ran a 280 trial block of each task, then a second block of each task in the same order, with a five-minute break after the second and fourth blocks.

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

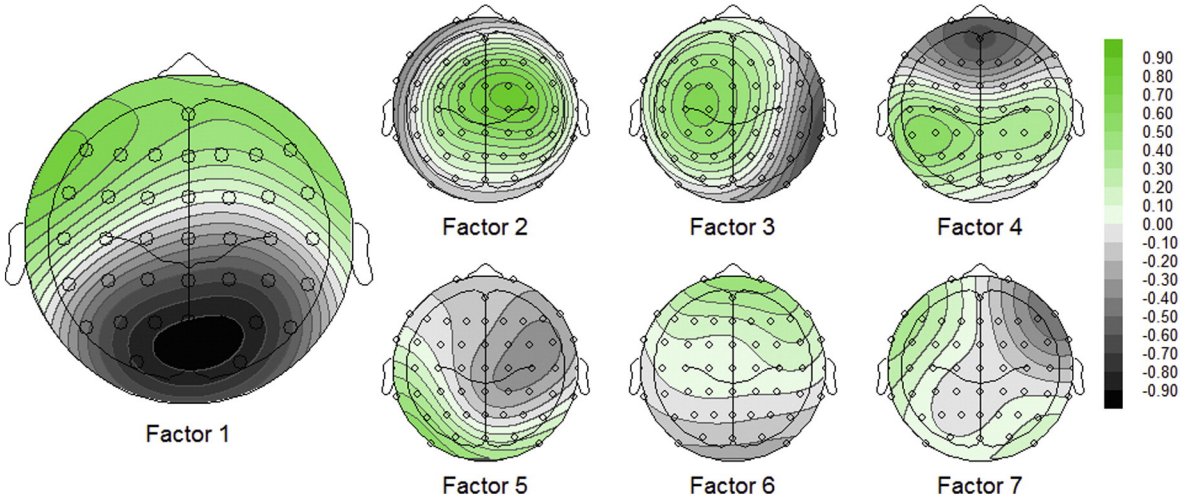

Fig. 1. Topographies for the seven factors yielded by Infomax-rotated Principal Components Analysis. Factor 1 used for all analyses.

Stimuli were white Arial letters (visual angle = ~3°) on a black background. Stimuli were presented for 1000 ms and stimulus-onset asynchrony varied uniformly between 1950 and 2050 ms in 25 ms steps. Stimuli were delivered using E-Prime V1.2 (Psychology Software Tools, PA, USA).

N-back tasks that were chosen as their oddball structure is more appropriate for eliciting prominent P3b components than the tasks more commonly used to assess the worst performance rule (Polich, 1987), which generally employ more balanced stimulus probabilities. Furthermore, in n-back tasks working memory load can be altered parametrically while using identically formatted stimuli.

3. Data analysis

3.1. EEG data

Data for individual participants were concatenated across blocks for each task separately. Periods where amplitude ranged by >1500 μV or <0.5 μV in any 200 ms window were rejected. Infomax independent components analysis (ICA) was then run on 180 s of data from each task, and factor weightings were applied to the entire task. Components reflecting oculomotor, cardiac, or electromyographic artefacts were identified by their time-courses and topographies and removed, then all remaining components were back-projected.

Data were common-average referenced and 0.05–50.00 Hz zero-phase Butterworth filters (24 dB/octave roll-offs) were applied. Remaining artefacts were removed by excluding data still ranging >150 μV in any 200 ms window, as well as data 200 ms before and after this section.

Data were re-filtered with 4 Hz low-pass Butterworth filters (24 dB/octave roll-offs) to improve signal-to-noise ratio prior to single-trial analysis (Smulders, Kenemans, & Kok, 1994). Oddball trials with a correct response were divided into segments ranging from 1400 ms pre-stimulus until 1650 ms post-stimulus. Data were baseline corrected using the −600 to −400 ms period. Shorter stimulus-locked (−600 to 1400 ms) and response-locked segments (−1000 to 400 ms) were then cut from the longer segments, so that they shared a common baseline.

Single-trial analysis was then carried out using the procedure from Saville et al. (2014, 2011a, 2012, 2015a,b), with modifications described below. We concatenated average stimulus-locked ERPs along the time-axis, and ran spatial principal components analysis on these data using the ERP-PCA toolkit (Dien, 2010). A parallel scree test found seven factors, which were extracted and Infomax-rotated (Bell & Sejnowski, 1995). Factor 1 had the P3b topography shown in Fig. 1, so was used hereafter. Single-trial stimulus-locked and response-locked data were summed with weightings derived from Factor 1's pattern factor matrix to produce virtual Factor 1 time-courses using the factor pattern matrix as a spatial filter.

In stimulus-locked trials of the factor 1 time-course, peaks were identified in each trial as the time-point 250–1000 ms post-stimulus with maximal amplitude.[1] In response-locked trials of the factor 1 time-course, peaks were identified in each trial as the time-point between −375 and +375 ms, relative to RT, with maximal amplitude. Trials where P3b latencies were identified at the edges of the window were excluded, as these latencies are likely to represent misidentified peaks, as were trials where response-locked latencies were identified before stimulus-onset.

Stimulus-locked and response-locked P3b latencies, and RTs, for all tasks and participants were then concatenated, along with information on participant identity, *g*, and task.

3.2. Inferential statistics

3.2.1. Worst performance rule analysis

To test for the worst performance rule in our data, we employed a novel quantile bootstrapping procedure. Implementing a standard case resampling bootstrap (Efron & Tibshirani, 1993), we randomly resampled, with replacement, each participant's single-trial P3b latencies. For each resampling, we divided the latencies into sixteen evenly sized quantile bins, and took the mean latency of each bin. A Spearman's rank correlation coefficient was then computed between participants' means for each quantile and their *g* scores. This procedure was repeated 10,000 times for each task, and values were computed for stimulus and response-locked P3b latency, as well as RT, data, separately.

In order to examine whether correlations varied as a function of quantile, the bootstrapped correlation coefficients were Fisher-transformed, averaged across all three tasks, and inverse Fisher-transformed back into correlation coefficients. Confidence intervals (CIs) for the correlation coefficients at each quantile were then derived from the bootstrapped data for stimulus and response-locked P3b data, and RT data, separately. By comparing CIs of correlation coefficients in the fastest quantiles to those in the slowest, we can formally test for evidence of a *worst performance rule*.

Presenting 95% confidence intervals is informative because readers are likely to have an intuitive sense for this criterion. However, this analysis makes multiple comparisons and so the risk of false discoveries is

[1] The sign of factors from Infomax rotation is arbitrary. Electrodes near Pz were weighted negatively so amplitude was given as a negative figure. This should not be confused with electrical polarity and, to avoid confusion, we refer to amplitude of the P3b factor as positive.

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

Table 1
Descriptive statistics for EEG and behavioural variables. SDs for each quantile in square brackets. Correlations are between SDs for each P3b and RT variable and proportions of accurate trials, and represent Spearman's *rho* values.

	Load	Mean of participant means	SD of participant means	Range of participant means	Correlation between variability and *g*	Mean quantile 1	Mean quantile 2	Mean quantile 3	Mean quantile 4	Mean quantile 5	Mean quantile 6
Stimulus-locked latency (ms)	0BT	473	43	379–568	−.16	322 [32]	358 [31]	376 [30]	390 [32]	402 [33]	413 [33]
	1BT	505	61	408–696	−.26	322 [35]	359 [36]	380 [38]	396 [41]	413 [47]	429 [54]
	2BT	536	65	428–756	−.17	301 [35]	349 [45]	382 [50]	409 [63]	429 [69]	453 [72]
Response-locked latency (ms)	0BT	−39	40	−113–35	−.14	−263 [56]	−171 [62]	−128 [50]	−103 [47]	−87 [46]	−73 [45]
	1BT	−31	47	−174–93	−.29	−272 [57]	−181 [65]	−140 [57]	−111 [52]	−90 [51]	−72 [48]
	2BT	−57	56	−216–81	−.17	−319 [45]	−250 [68]	−202 [74]	−159 [71]	−131 [69]	−105 [65]
Reaction time (ms)	0BT	492	48	393–600	.12	379 [37]	413 [41]	430 [42]	442 [44]	453 [46]	464 [46]
	1BT	513	55	386–636	−.12	369 [42]	403 [45]	422 [48]	440 [51]	454 [51]	469 [53]
	2BT	599	70	414–730	−.02	389 [47]	431 [56]	460 [63]	484 [70]	505 [72]	526 [73]
Accuracy (% correct)	0BT	92	5	77–100	.02	–	–	–	–	–	–
	1BT	88	6	72–98	.10	–	–	–	–	–	–
	2BT	74	12	44–93	.32	–	–	–	–	–	–

Table 1 (*continued*)

	Mean quantile 7	Mean quantile 8	Mean quantile 9	Mean quantile 10	Mean quantile 11	Mean quantile 12	Mean quantile 13	Mean quantile 14	Mean quantile 15	Mean quantile 16
Stimulus-locked latency (ms)	423 [33]	435 [34]	447 [37]	462 [42]	480 [48]	502 [57]	532 [72]	575 [95]	646 [122]	795 [123]
	445 [60]	462 [64]	480 [69]	502 [77]	526 [84]	556 [90]	593 [100]	644 [107]	722 [121]	855 [99]
	477 [77]	500 [81]	522 [84]	548 [87]	578 [86]	610 [91]	649 [99]	696 [102]	770 [92]	895 [67]
Response-locked latency (ms)	−61 [43]	−47 [42]	−36 [42]	−25 [42]	−11 [42]	6 [44]	26 [54]	56 [72]	99 [92]	192 [98]
	−55 [48]	−41 [47]	−27 [48]	−13 [50]	5 [55]	22 [59]	43 [63]	75 [73]	126 [86]	233 [83]
	−85 [64]	−63 [61]	−43 [66]	−24 [68]	−7 [72]	14 [71]	36 [75]	69 [82]	120 [93]	228 [91]
Reaction time (ms)	473 [48]	483 [49]	492 [50]	502 [51]	511 [52]	523 [52]	536 [54]	552 [55]	577 [60]	649 [81]
	482 [54]	496 [54]	509 [55]	525 [56]	540 [58]	556 [60]	574 [62]	600 [67]	635 [72]	731 [100]
	548 [75]	569 [76]	590 [77]	613 [79]	638 [80]	665 [83]	697 [88]	739 [91]	795 [90]	926 [98]
Accuracy (% correct)	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–

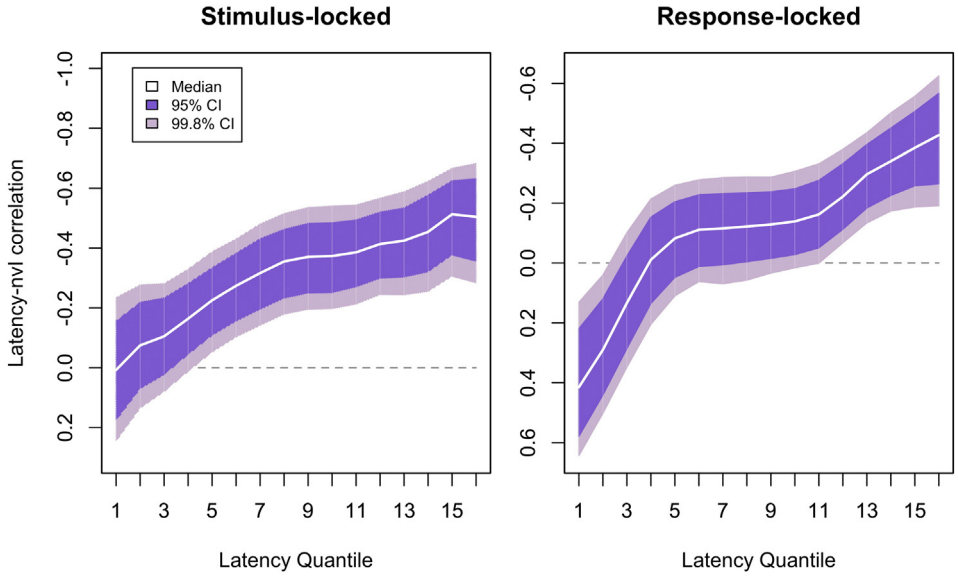

Fig. 2. Confidence intervals for bootstrapped Spearman's correlation coefficients between latency and *g* for each quantile of the P3b latency distribution separately. Coefficients were transformed into Fisher's Z, averaged across tasks, and re-transformed into *r*. Data for stimulus and response-locked data presented separately.

elevated. Hence we also present additional 99.8% CIs, a stricter criterion equivalent to Bonferroni-correcting for comparing the two fastest quantiles with the two slowest quantiles in stimulus and response-locked P3b data, and in RT data ($4 * 3 = 12$ comparisons). This should help readers make informed judgements about the robustness of our findings. In the interest of transparency, it should be noted that an earlier version of our analysis employed only six quantiles, but on the advice of reviewers we switched to the 16 used by Larson and Alderton (1990), which offer both greater granularity and comparability to previous data. The general pattern of results is unchanged. The comparison of the top two and bottom two quantiles was retained for formal hypothesis testing.

3.2.2. Linear mixed effects model

To verify that stimulus and response-locked measures of P3b latency made independent contributions to RTs, we fitted linear mixed effects models to the PCA-weighted single-trial data using the *lme4* package (Bates, Maechler, & Bolker, 2012) for R (R Core Development Team, 2012). Our full model predicted RT, with fixed effects of SL-P3B (stimulus-locked P3b latency), RL-P3B (response-locked P3b latency), and LOAD (0BT, 1BT, and 2BT, coded as a linear variable), and random intercepts and slopes of each fixed effect for each participant separately. This full model was compared to reduced models, dropping the fixed effects of SL-P3B, RL-P3B, or both, using Bayesian information criteria. All models were fitted using log-likelihood estimation.

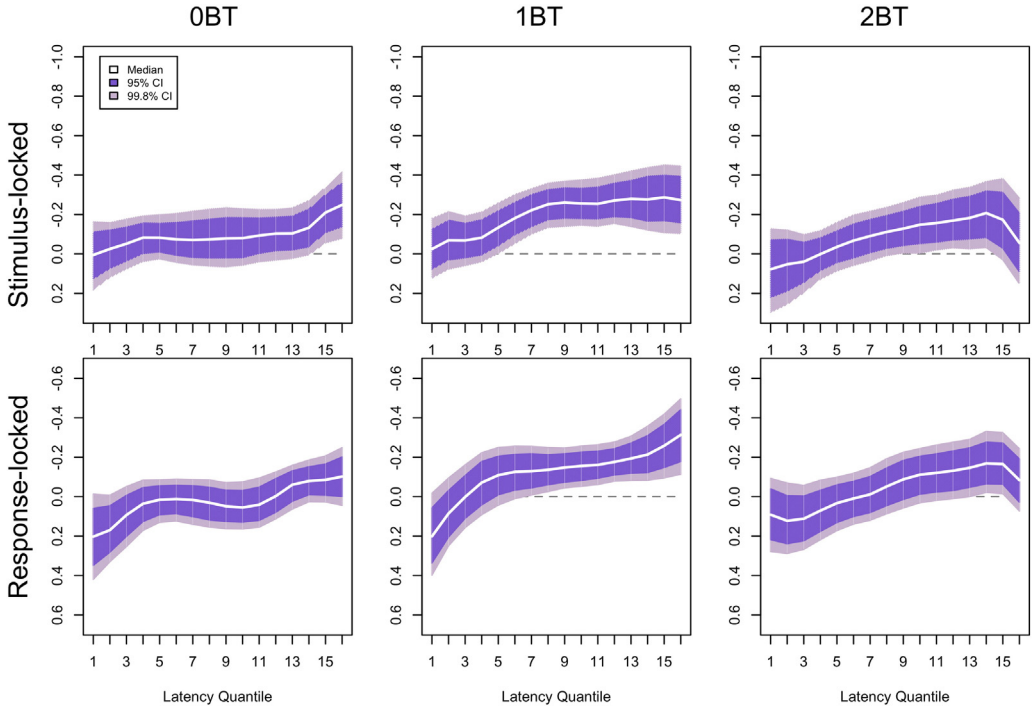

Fig. 3. Confidence intervals for bootstrapped Spearman's correlation coefficients between latency and *g* for each quantile of the P3b latency distribution separately. Data for different tasks presented separately for stimulus and response-locked data. These data were not used for hypothesis testing, but are presented to help illustrate our findings.

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

We also tested the best-performing model against versions including G, and G with its interaction with SL-P3B, RL-P3B, and both SL-P3B and RL-P3B.

3.3. Behavioural data

The quantile bootstrapping analysis described above was also run using RTs associated with the trials used in the P3b analysis. CIs at the 95% and 99.8 level were compared across quantiles to assess whether RT data showed evidence of a worst performance rule.

4. Results

Table 1 shows descriptive statistics for stimulus and response-locked P3b latencies; RT; and accuracy.

4.1. EEG data

4.1.1. Worst performance rule analysis

Fig. 2 shows bootstrapped correlation coefficients at each quantile of the stimulus and response-locked P3b latency distributions. The results for stimulus-locked data follow a clear worst performance rule pattern. Even at the highly stringent 99.8% criterion, confidence intervals: a) do not include zero, after the first four quantiles, and b) become more negative with increasing quantile, so that faster latencies in these later quantiles are associated with higher *g*.

This can be verified statistically by showing that in stimulus-locked data the confidence intervals for the first two and last two quantiles do not overlap using the 99.8% CIs. This analysis shows strong evidence for a worst performance rule in stimulus-locked P3b data.

The pattern is more complicated in the response-locked data. While correlations show the same slope as in stimulus-locked data, non-zero correlation coefficients of opposite signs can be seen in both tails of the distribution, reflecting that high *g* is associated with a) fewer P3b latencies which trail the response by a long latency, b) fewer P3b latencies which precede the response by a long latency. In short, low *g* participants have response-locked distributions that are broader than those with high *g*. Statistically, the 99.8% CIs for the first two quantiles, once again, do not overlap with the last two quantiles.

Thus, although both stimulus and response-locked P3b latencies show variable associations with *g* at different quantiles, the presence of non-zero opposite signed effects in the two tails of the response-locked distribution is a qualitatively different pattern to the *worst performance rule* present in stimulus-locked data.

Although hypothesis testing was carried out using correlation coefficients that had been averaged across tasks, to improve the robustness of our results by means of data aggregation, we present plots of bootstrapped correlation coefficients for all tasks separately in Fig. 3. Patterns found in individual tasks resemble those found in the averaged data, suggesting that the figures presented here are representative and appropriate for statistical analysis. The pattern appears to be clearest in the 1BT.

For illustration purposes, we also divided our participants into two groups, based on a median-split of *g* scores. Quantile means at each iteration of the bootstrap were averaged within group into quantile grand means. In Fig. 4, we present 95% CIs of stimulus and response-locked data for the two groups on each task separately. Our hypothesis testing was carried out on the correlation data, but we include these data for interested readers.

4.1.2. Linear mixed effects analysis

The best-fitting model to predict RTs included SL-P3B, RL-P3B, and LOAD, but not G or its interactions (BIC = 177,141.9). This outperformed the models dropping SL-P3B (BIC = 177,217.1), RL-P3B (BIC = 177,240.5), and LOAD (BIC = 177,193.4), as well as the models adding G (BIC = 177,149.0), G and its interaction with SL-P3B (BIC = 177,153.8), G and its interaction with RL-P3B (BIC = 177,153.2), and G and its interactions with both P3b variables (BIC = 177,162.0).

The chosen model showed that RT was positively associated with SL-P3B ($\beta = .44$, $\sigma_\beta = .03$, $t = 13.4$), and LOAD ($\beta = 34.1$, $\sigma_\beta = 3.4$, $t = 10.0$), and negatively associated with RL-P3B ($\beta = -.46$, $\sigma_\beta = .04$, $t = -12.9$; intercept: $\beta = 224.3$, $\sigma_\beta = 13.4$, $t = 16.7$). As both P3b latency and RT are in the same units (ms), the β-slopes for SL-P3B and RL-P3B can be interpreted similarly to R^2 values.

4.2. Behavioural data

Fig 5 shows bootstrapped correlation coefficients for RT data. While there is some evidence of increasingly strong correlations between RT and *g* from the earliest quantiles to the middle quantiles, this pattern reverses in the later quantiles, and the pattern appears qualitatively different to a worst performance rule. All confidence intervals substantially

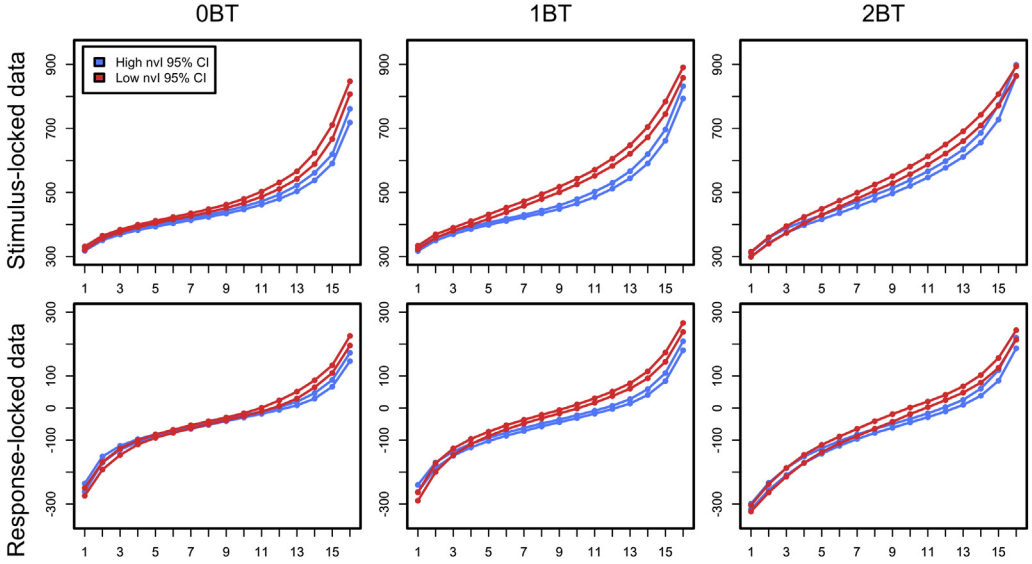

Fig. 4. Bootstrapped mean latencies for each quantile of the P3b latency distribution separately. Blue lines represent participants scoring above the median *g*, red lines represent those scoring below the median. These data were not used for hypothesis testing, but are presented to help illustrate our findings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

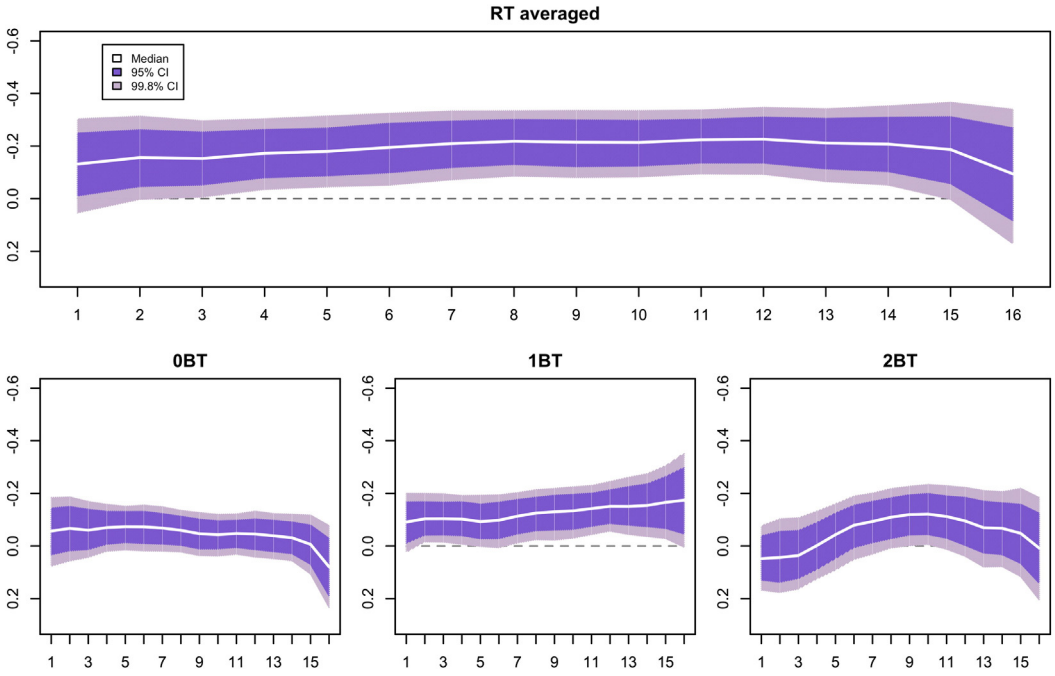

Fig. 5. Confidence intervals for bootstrapped Spearman's correlation coefficients between RT and *g* for each quantile of the RT distribution separately. Top panel shows coefficients aggregated across tasks (data were transformed into Fisher's Z, averaged across tasks, and re-transformed into *r*). Bottom three panels represent data from individual tasks. The latter data were not used for hypothesis-testing but are presented for illustrative purposes.

overlap, and so the null hypothesis cannot be rejected. As with the P3b latency data, although hypothesis tests were carried out on correlation coefficients that had been averaged across tasks, we also include plots of the quantile-wise correlation coefficients for all three tasks separately in Supplementary Fig. 6. Unlike those for P3b data, the plots for the separate tasks are heterogeneous across tasks, reinforcing our impression that a worst performance rule pattern is not present in our RT data. While the data for the 1BT are qualitatively in line with a worst performance rule, although CIs still overlap, the patterns for the 0BT and 2BT are not. The 0BT, if anything, shows the opposite trend while the 2BT shows an increase in correlation magnitude from low to middle quantiles, followed by a drop from middle to late quantiles. Interestingly these parallel the results of the overall correlations between these variables and *g* estimates, shown in Table 1 — 0BT RT variability shows an unusual positive correlation with *g*, while the 2BT RTs show essentially no correlation, and the 1BT RTs trend weakly in the normal direction.

That said, it is important to note that confidence intervals overlap for all tasks at each quantile, so while the patterns appear different for the three tasks, caution should be taken not to overinterpret these differences.

5. Discussion

We examined the association between the shape of the P3b latency distribution and psychometric intelligence, and identified quantile-dependent associations with intelligence in both stimulus and response-locked P3bs. Stimulus-locked data showed evidence of a worst performance rule pattern, while response-locked data showed that lower *g* was associated with a generally broader distribution of P3b latencies. However, like Salthouse (1998) but unlike other authors (Coyle, 2003; Kranzler, 1992; Larson & Alderton, 1990), we did not observe a worst performance rule pattern in the RT data themselves.

It is important to consider why a worst performance rule was not found in our behavioural data. One possibility is that this is due to the dual-task nature of the tasks we used — participants had to both update working memory and respond to matches and this may have contaminated RT as a measure of processing speed while leaving P3b latency relatively uncontaminated. It should be noted, however, that the 0BT was not a dual task and, if anything, showed the greatest deviation from the worst performance rule.

A second possibility is that the failure to find a worst performance rule is simply a Type II error. That said, although the present study has a smaller sample than previous behavioural studies (Kranzler, 1992; Larson & Alderton, 1990) due to the more arduous process of data collection in EEG studies, and thus less statistical power, power calculations suggest that the present study still had a power of ~75% to detect the effect sizes identified in the previous studies ($r = \sim.37$) assuming a two-tailed hypothesis, and an ~85% power assuming a directed hypothesis, suggesting that it was not obviously underpowered. Our procedure of comparing quantile correlations to one another was arguably a more stringent test than comparing correlation magnitudes to zero, the procedure used in previous studies, but, with the exception of the 1BT, the RT data do not even trend towards a worst performance rule pattern. Thus while it remains a possibility which future replications will need to address, this is not obviously a Type II error.

A third possibility is that the EEG measures may be more proximal to the underlying phenomenon than RT measures and are thus more sensitive. This line of reasoning lies behind the concept of the endophenotype, or intermediate phenotype, in psychiatry (Gottesman, & Gould, 2003), where genetic associations with neural measures have been shown to have greater effect sizes than with behavioural measures, at least in the context of psychosis risk genetics (Rose & Donohoe, 2013). It is not clear whether this reflects publication bias, the greater 'researcher degrees of freedom' associated with neural measures, or the greater proximity that is the logic behind the endophenotype concept. Also along this line, our data could be seen as converging with Jensen's (1998) demonstration that *g* is more highly correlated with selective measures of RT than measures of movement time. P3b latency is thought to be relatively sensitive to the same sort of factors that influence RT, but relatively insensitive to those that influence movement time. However, we acknowledge that our RT measure does not separate RT from movement time, which might have

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

impaired the replication of the worst performance rule at the behavioural level.

A fourth possibility is that the 'both-tails rule' found in the response-locked data masked the worst performance rule in the RT data. This could be due to a separate relationship between response-planning reliability and *g*, or could reflect the complexity of interpreting dual task data described above.

Fifthly, it is worth pointing out that Salthouse (1998) also did not find a worst performance rule pattern, so it is possible that the effect only manifests under certain circumstances. As mentioned above, RTs in the 1BT results trended in the direction of a worst performance rule (Fig. 5) and the neural analogue for the worst performance rule was also clearest in this task (Fig. 3). This could suggest that the worst performance rule is most prominent with an intermediate level of difficulty; the 0BT may have been too easy and the 2BT may have been too difficult. However Kranzler (1992) reported a worst performance rule in choice RTs, which are very similar to the 0BT reported here. Furthermore, confidence intervals overlapped for RTs in all three tasks at every quantile. We are thus wary of overinterpreting these apparent cross-task differences.

Finally, to continue the previous point, our study design was optimised to detect P3b effects. For example, we used an oddball design to improve signal-to-noise ratio and we did not use an apparatus that separates RT and movement time to minimise participant movement. Optimal conditions for a P3b effect, however, may not be optimal for detecting a behavioural effect. In light of this, we are wary of drawing strong conclusions about the robustness of the behavioural worst performance rule based on our data.

This lack of a behavioural effect suggests that our EEG findings should not be viewed as the neural *correlates* of the worst performance rule, but instead as an interesting EEG *analogue* to this effect that exists in its own right. As Wilkinson and Halligan (2004) argue, although finding both neural and behavioural effects certainly makes for simpler interpretations, neural effects without a corresponding behavioural finding can still provide important insights.

A worst performance rule was, however, found in stimulus-locked P3b latency data, and a novel quantile-dependent association between *g* and P3b latency was found in response-locked P3b data. The present findings extend previous results linking P3b latency and *g* (e.g. De Pascalis, Varriale, & Matteoli, 2008), and identify possible relationships between *g* and different processing stages. That the stimulus-locked data show such a pattern suggests that it could reflect an association between psychometric intelligence and pre-decision processes. This is consistent with both the attentional lapse and diffusion models of the worst performance rule. However our identification of a 'both tails' pattern in response-locked data suggests that the association between *g* and the speed of information processing is not just due to pre-decision processes. Individual differences in *g* appear to also be associated with later stages of processing, perhaps motor planning and execution, and low scorers exhibit more variability in these stages of processing as well. Such associations may be less apparent using purely behavioural measures, and so their identification with single-trial EEG is a key novel finding of the present study.

It is unclear whether the patterns seen in stimulus-locked and response-locked data have the same source, or whether separate sources of variability underlie these two distributions. Although previous work has shown that stimulus and response-locked P3b latency variability are highly correlated in healthy young participants (Saville et al., 2012), there is also evidence for response-locked variability being specifically increased in attention-deficit hyperactivity disorder (Saville et al., 2015b), suggesting that the two are dissociable.

In the present data, attentional lapses causing late stimulus-locked P3b latencies could sometimes be offset by parallel processes leading to the RT preceding P3b. This, accompanied by uncompensated-for lapses, represented by long stimulus and response-locked P3b latencies, could cause this symmetrically broader response-locked P3b distribution in low scorers. Likewise, a domain general process could lead to noisier cognition in both pre-decision and post-decision processing, explaining both findings within the diffusion model framework. On the other hand, the 'both-tails rule' could represent an entirely distinct phenomenon to the worst performance rule, related to specific problems with response-planning. Further work will make this clear.

Finally, it is important to point out possible technical ambiguities in our findings. As P3b latency must be measured from a noisy signal, a proportion of the identified peaks will not reflect the true P3b latency. A risk with this approach is that misidentified peaks may not occur uniformly through the distribution, thus confounding analyses of specific section of the distribution. We excluded trials where latency was recorded at the borders of the peak picking window as a precaution against this, but if the latency distribution of the P3b overlapped with that of another deflection with a similar topography, this could lead to a greater incidence of peak misidentification in certain sections of the distribution.

To conclude, we found evidence for a worst performance rule, analogous to that seen in RT data in previous studies, in single-trial stimulus-locked P3b latency data. We also found evidence for a symmetrically broader response-locked P3b latency distribution in low scorers — a *both tails 'rule'*. However we did not observe the worst performance rule in the RT data, and so care must be taken in interpreting these findings. Single-trial ERPs are an important technique for measuring distributional changes in the timing of neurocognitive operations and mental events, offering great promise for exploring the worst performance rule and the neural basis of *g*.

Acknowledgements

The authors would like to thank two anonymous reviewers for their helpful and constructive comments on the manuscript.

References

American Electroencephalographic Society (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *2*, 200–202.
Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Mixed-effects modeling with R.*
Beauducel, A., Liepmann, D., Horn, S., & Brocke, B. (2010). *Intelligence structure test. English version of the Intelligenz-Struktur-Test 2000 R (IST 2000 R).* Göttingen: Hogrefe.
Beck, L. (1933). The role of speed in intelligence. *Psychological Bulletin*, *4*, 169–178 (Retrieved from http://psycnet.apa.org/journals/bul/30/2/169/).
Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.
Coyle, T. R. (2003). IQ, the worst performance rule, and Spearman's law: A reanalysis and extension. *Intelligence*, *31*(5), 473–489. http://dx.doi.org/10.1016/S0160-2896(02)00175-7.
Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study*29*, 389–399.
De Pascalis, V., Varriale, V., & Matteoli, A. (2008). Intelligence and P3 components of the event-related potential elicited during an auditory discrimination task with masking. *Intelligence*, *36*(1), 35–47.
Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145. http://dx.doi.org/10.1016/j.jneumeth.2009.12.009.
Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, *57*.
Gerson, A. D., Parra, L. C., & Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage*, *28*(2), 342–353. http://dx.doi.org/10.1016/j.neuroimage.2005.06.026.
Gottesman, I. I., & Gould, T. D. (2003). *Reviews and overviews the endophenotype concept in psychiatry : Etymology and strategic intentions.* , 636–645 (April).
Jensen, A. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, *13*(8), 869–881.
Jensen, A. (1998). *The g factor: The science of mental ability.* Santa Barbara: Praeger.
Jensen, A. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier (Retrieved from http://books.google.com/books?hl=en&lr=&id=0EBypqsH-0IC&oi=fnd&pg=PP2&dq=Clocking+the+mind:+Mental+chronometry+and+individual+differences&ots=RgA5KcShES&sig=ubxgATngm32oZD5isOo72sNkLR8).
Kranzler, J. (1992). A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences*, *13*(3), 255–261 (Retrieved from http://www.sciencedirect.com/science/article/pii/019188699290099B).
Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: Thse P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795 (Retrieved from http://www.sciencemag.org/content/197/4305/792.short).

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309–325. http://dx.doi.org/10.1016/0160-2896(90)90021-K.

Miller, E. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, *8869*(94), 803–832 (Retrieved from http://www.sciencedirect.com/science/article/pii/0191886994900493).

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. http://dx.doi.org/10.1037/0033-2909.131.4.510.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735 (http://doi.org/10.1038/nn.3248).

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, *68*(4), 311–320. http://dx.doi.org/10.1016/0168-5597(87)90052–9.

R Core Development Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*(1), 10–17. http://dx.doi.org/10.1016/j.intell.2006.12.002.

Rose, E. J., & Donohoe, G. (2013). Brain vs behavior: an effect size comparison of neuroimaging and cognitive studies of genetic risk for schizophrenia. *Schizophrenia Bulletin*, *39*(3), 518–526. http://dx.doi.org/10.1093/schbul/sbs056.

Salthouse, T. A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence*, *26*(2), 153–166. http://dx.doi.org/10.1016/S0160-2896(99)80059–2.

Saville, C. W. N., Feige, B., Kluckert, C., Bender, S., Biscaldi, M., Berger, A., ... Klein, C. (2015a). Increased reaction time variability in attention-deficit hyperactivity disorder as a response-related phenomenon : Evidence from single-trial event-related potentials. *Journal of Child Psychology and Psychiatry*, *56*(7), 801–813. http://dx.doi.org/10.1111/jcpp.12348.

Saville, C. W. N., Lancaster, T. M., Davies, T. J., Toumaian, M., Pappa, E., Fish, S., ... Klein, C. (2015b). Elevated P3b latency variability in carriers of ZNF804A risk allele for psychosis. *NeuroImage*, *116*, 207–213. http://dx.doi.org/10.1016/j.neuroimage.2015.04.024.

Saville, C. W. N., Dean, R. O., Daley, D., Intriligator, J., Boehm, S., Feige, B., & Klein, C. (2011a). Electrocortical correlates of intra-subject variability in reaction times: Average and single-trial analyses. *Biological Psychology*, *87*(1), 74–83. http://dx.doi.org/10.1016/j.biopsycho.2011.02.005.

Saville, C. W. N., Lancaster, T. M., Stefanou, M. E., Salunkhe, G., Lourmpa, I., Nadkarni, A., ... Klein, C. (2014). COMT Val158Met genotype is associated with fluctuations in working memory performance: Converging evidence from behavioural and single-trial P3b measures. *NeuroImage*, *100*, 489–497. http://dx.doi.org/10.1016/j.neuroimage.2014.06.006.

Saville, C. W. N., Pawling, R., Trullinger, M., Daley, D., Intriligator, J., & Klein, C. (2011b). On the stability of instability: Optimising the reliability of intra-subject variability of reaction times. *Personality and Individual Differences*, *51*(2), 148–153. http://dx.doi.org/10.1016/j.paid.2011.03.034.

Saville, C. W. N., Shikhare, S., Iyengar, S., Daley, D., Intriligator, J., Boehm, S. G., ... Klein, C. (2012). Is reaction time variability consistent across sensory modalities? Insights from latent variable analysis of single-trial P3b latencies. *Biological Psychology*, *91*(2), 275–282. http://dx.doi.org/10.1016/j.biopsycho.2012.07.006.

Smulders, F. T. Y., Kenemans, J. L., & Kok, A. (1994). A comparison of different methods for estimating single-trial P300 latencies. *Electroencephalography and Clinical Neurophysiology*, *92*(2), 107–114. http://dx.doi.org/10.1016/0168-5597(94)90051–5.

Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *20*(2), 165–181. http://dx.doi.org/10.1027/0269-8803.19.2.xxx.

Verleger, R., Schroll, H., & Hamker, F. (2013). The unstable bridge from stimulus processing to correct responding in Parkinson's disease. *Neuropsychologia*, *51*, 2512–2525 (Retrieved from http://europepmc.orssg/abstract/med/24051004).

Wilkinson, D., & Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews Neuroscience*, *5*(1), 67–73. http://dx.doi.org/10.1038/nrn1302.

Please cite this article as: Saville, C.W.N., et al., A neural analogue of the worst performance rule: Insights from single-trial event-related potentials, *Intelligence* (2015), http://dx.doi.org/10.1016/j.intell.2015.12.005